Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Public Health Theses | School of Public Health

1-1-2019

Multi-State Analysis Of Unconventional Oil And Gas Development And Rates Of Sexually Transmitted Infections In The United States

Nicholaus Peter Johnson
nick.peter.j@gmail.com

Multi-State Analysis of Unconventional Oil and Gas Development and Rates of Sexually Transmitted Infections in the United States

Nicholaus Johnson

Thesis Advisor & First Reader: Dr. Nicole Deziel

Second Reader: Dr. Linda Niccolai

Collaborator: Dr. Joshua Warren

Yale School of Public Health

April 15, 2019

Table of Contents

List of Figures

List of Tables

Abstract

Background

Advances in unconventional oil and gas development (UOGD), including hydraulic fracturing and horizontal drilling, have helped position the United States as a world leader in the production of both crude oil and natural gas hydrocarbons. This significant increase in production would not have been possible, however, without the assemblage of a workforce that is predominantly young, male, and mobile. This is significant given that a number of studies have associated similar types of mobile workforces – especially those employed by resource extractive industries – with impacts on the sexual health of the host communities in which these workers operate (Westwood & Orenstein, 2016; Udoh et al., 2008). For instance, two ecologic studies have explored the association between shale activity and county level rates of reportable STI in counties overlying the Marcellus Shale, with positive associations being reported in each. While their findings are novel, these studies are restricted in their geographic scope. Some of the largest and most productive oil and gas formations are located in other regions of the United States including Colorado, North Dakota, and Texas. A multi-state, multi-region analysis of the association between shale activity and county level rates of reportable STI – chlamydia, gonorrhea, and primary/secondary syphilis – is needed to confirm whether this previously observed association is a unique place-based phenomenon or a generalizable result of UOGD.

Methods

To evaluate the association between shale drilling activity and rates of STI in Colorado, North Dakota, and Texas, we conducted an ecologic study that utilized annual census, STI, and drilling activity data at the county level from 2000-2016. We requested the annual rates (cases per 100,000 population) of chlamydia, gonorrhea, and primary/secondary syphilis for the 64 counties in Colorado, 53 counties in North Dakota, and 254 counties in Texas using the Centers for Disease Control and Prevention's (CDC) National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP) Atlas Plus tool. Data pertaining to drilling activity was collected annually for each county through DrillingInfo's DI Application. Each county year was classified according to its shale activity: none (0 spuds targeting shale), low (1-49 spuds targeting shale), and high (≥ 50

spuds targeting shale). Data for sociodemographic covariates was obtained from the Decennial Census and American Community Survey (ACS) provided by the U.S. Census Bureau's American Fact Finder tool. We used multivariable mixed effects Poisson regression models to estimate rate ratios (RR) with 95% CI for the association between county year shale activity and cases of chlamydia, gonorrhea, and primary/secondary syphilis for each state, while adjusting for potential confounders and secular trends.

Results

Compared to county years with no shale activity, Texas county years with high shale activity (≥ 50 shale spuds) had 10% (RR = 1.10; 95% CI = 1.04, 1.17) increased rates of chlamydia and 15% (RR = 1.15; 95% CI = 1.04, 1.28) increased rates of gonorrhea. No association was observed for syphilis in Texas. An association between shale activity and annually reported rates of chlamydia, gonorrhea, and syphilis were not observed for Colorado and North Dakota.

Conclusions

The first multi-state, multi-region analysis of shale activity and annual STI rates confirmed previously observed associations in Texas counties, for the association between shale drilling activity and annual rates of chlamydia and gonorrhea. While elevated rates in the unadjusted models in Colorado and North Dakota suggested a positive association, these effect estimates did not remain after rigorous adjustments for sociodemographic covariates, secular trends, county-level random effects, and observation-level random effects.

Introduction

In recent years the United States has become a world leader in the production of both crude oil and natural gas hydrocarbons. In 2018 the United States Energy Information Administration (EIA) estimated that the United States had surpassed both Saudi Arabia and Russia as the world's largest producer of crude oil (Dunn & Hess, 2018). In addition to petroleum, the United States was also the world's largest producer of natural gas hydrocarbons between 2008 and 2017 (EIA, 2018). These increases were made possible in part by the rapid adoption and deployment of new forms of Unconventional Oil and Gas Development (UOGD), including hydraulic fracturing and horizontal drilling. For instance, in 2014 it was estimated that 95% of wells drilled that year were hydraulically fractured and that shale gas production would be almost completely responsible for significant future increases in natural gas production (Moniz, 2014). Indeed, by 2015 hydraulically fractured wells were responsible for two-thirds of natural gas production (Perrin & Cook, 2016) and more than 50% of crude oil production in the U.S. (Cook & Perrin, 2016).

This recent energy production boom is not without potential deleterious effects on public health, especially in those communities with the highest levels of activity. UOGD has been associated with the deterioration of several markers of community health including water quality (Elliott et al., 2015), air quality (Litovitz et al., 2013), mental health (Hirsch et al., 2018), social cohesion (Morrone et al., 2015), and noise levels (Hays et al., 2017).

Recently, UOGD has also been associated with an increase in certain STI throughout the northeastern U.S. For instance, one study found that in counties overlying the Marcellus Shale formation – including counties in Maryland, New Jersey, New York, Ohio, Pennsylvania, Virginia, and West Virginia – counties experiencing a "fracking boom" had a 20% increase in gonorrhea compared to counties not experiencing a "fracking boom" (Komarek & Cseh, 2017). Furthermore, a second study found that Ohio counties with high shale gas activity had 21% increased rates of chlamydia and 19% increased rates of gonorrhea when compared to counties throughout the state with no such activity (Deziel et al., 2018). It is suggested in these studies that the increase in community-level STI rates is likely attributable to the influx of workers needed to support drilling activity that is most pronounced at the beginning of the drilling process. This is consistent with a number of studies that have demonstrated a similar association between mobile workers – especially those employed by resource extractive industries – and the sexual health of the host communities (Westwood & Orenstein, 2016; Udoh et al, 2008).

While these studies and their results are novel, they are restricted in their geographic scope. Both studies focused exclusively on communities in the Marcellus Shale region of the eastern United States. Additional studies are needed to understand whether these patterns are also observed in other regions, including in the midwestern and western United States. Colorado, North Dakota, and Texas were therefore chosen for this study as they are home to highly productive shale formations in the U.S. in terms of both crude oil and natural gas production (See Appendix A, Figures 4 and 5). The purpose of this study is to examine the relationship between shale drilling activity and rates of chlamydia, gonorrhea, and syphilis in Colorado, North Dakota, and Texas from 2000-2016. By doing so, this multi-state, multi-region analysis will investigate whether this previously observed relationship is a unique place-based phenomenon or a generalizable result of UOGD that should be taken into consideration when developing public health and energy policies throughout the U.S.

Methods

Study Design

To evaluate the association between shale drilling activity and rates of STI in Colorado, North Dakota, and Texas we designed an ecologic study that utilized annual census, STI, and drilling activity data at the county level from 2000-2016. This time period allows for the collection and incorporation of baseline census, STI, and drilling activity data prior to shale drilling activity, which commenced in either 2005 or 2006 depending on the state (See Appendix A, Figure 6). The following data were collected and organized annually for each county to allow for the future use of county years as the unit of analysis. Each county therefore represented 17 observations, one for each of the 17 years in the study period (2000-2016).

Data Sources

Sexually transmitted infections. We requested the annual rates (cases per 100,000 population) of chlamydia, gonorrhea, and syphilis which were collected for the 64 counties in Colorado, 53 counties in North Dakota, and 254 counties in Texas for the years 2000-2016 using the Centers for Disease Control and Prevention's (CDC) National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP) Atlas Plus tool (CDC, 2018a). Case data used to calculate these rates are confirmed diagnoses that are reported to the CDC by county and state health departments and population denominators are based on calendar year estimates from the U.S. Census Bureau (CDC, n.d.).

Sociodemographic covariates. A county's STI rate is known to be affected by a number of covariates including race/ethnicity, sex, age, income, educational attainment, and access to quality health care (CDC, 2018b). Data for sociodemographic covariates was obtained from the Decennial Census and American Community Survey (ACS) provided by the U.S. Census Bureau's American Fact Finder tool (U.S. Census Bureau, 2017). The covariates for which data was collected include: sex (percent population male and percent population female), age (percent population 15-29 years of age), race (percent population identifying as white, black or African American, American Indian or Alaska Native, and Asian), ethnicity (percent population identifying as Hispanic or Latino), educational attainment (percent population 25 years of age and older that graduated high school and percent population 25 years of age and older with a Bachelor's degree), income (median household income in USD per year), poverty (percent population in poverty within the last 12 months), and health care access (percent of the population with health insurance coverage). Population density was calculated by dividing the population estimates used by the CDC by county area (in miles squared) as reported by the U.S. Census Bureau in 2010 (U.S. Census Bureau, 2017; See Appendix B, Table 2).

The availability of this sociodemographic data varied across the study period (U.S. Census Bureau, 2017; See Appendix B, Tables 1 and 2). Data from the 2000 Decennial Census was applied to the years 2000-2004. Data from the 2005-2009 ACS (5-year estimates) were applied to the years 2005-2009. The ACS (5-year estimates) ending in the given year were then applied to the years 2010-2016 (e.g. the 2011-2015 ACS 5-year estimates were applied to the year 2015). Since percent insured data only became available with the 2008-2012 ACS (5-year estimates), values for this covariate were assigned to the years 2000-2012.

Well / drilling activity. Data pertaining to drilling activity was collected through DrillingInfo's DI Application (DrillingInfo, 2019). To minimize the variability inherent in obtaining information from various local and state sources, we used this commercially available dataset that has been used previously to assess the relationship between shale activity and county level rates of gonorrhea (Komarek & Cseh, 2017). The well data for each state was restricted using the DI Application to only those wells with a first production date between January 1, 2000 and December 31, 2016. After download, the wells were further refined by keeping only those wells with an active status and a spud date between January 1, 2000 and December 31, 2016. The spud date is commonly used in the oil and gas drilling industries to refer to the specific date at which well drilling is initialized. Spud dates were used instead of other possible variables including permit date and first production date because it is likely a better measure of the beginning time period of well drilling and development that is indicative of a greater influx of workers and activity. Lastly, the wells were refined by including only those targeting a shale formation. Wells targeting a non-shale oil or gas formation existed in each state, but were excluded due to a desire in this study to focus specifically on the effects of UOGD. A well was identified as targeting a shale formation if the name of the target formation could be identified on the EIA's map of shale gas and oil plays or if the target formation name included the word "shale" (EIA, 2016; See Appendix A, Figure 7).This dataset was not limited to either production type – including CO_2, dry hole, gas, gas or coalbed, oil, oil or gas, other, and unknown – or drill type – including directional, horizontal, unknown, and vertical. For the study period the number of active spuds targeting a shale formation for all counties totaled 6,543 in Colorado, 12,642 in North Dakota, and 32,998 in Texas (See Appendix B, Table 3).

Shale drilling activity was quantified according to frequency of spuds targeting a shale formation per county year. Upon a visual inspection of the distribution of non-zero values for this variable (TAS_Targeting_Shale) across all three states, 50 spuds targeting a shale formation per county year appeared to be an appropriate delineator between high and low shale drilling activity (See Appendix A, Figures 8-10). Due to the high variability in frequency of spuds targeting shale per county year, the median did not appear to be an appropriate delineator. Each county year was then categorized into a multilevel variable (Cat_TAS_Targeting_Shale) according to its shale drilling activity: none (0 spuds targeting shale), low (1-49 spuds targeting shale), and high (≥ 50 spuds targeting shale).

Statistical Analysis

We first conducted descriptive analyses for the exposure, sociodemographic, and outcome variables in which the mean, standard deviation, and analysis of variance (ANOVA) was calculated according to the categorization of county years by the shale drilling activity categories mentioned above (See Appendix B, Table 4). In bivariable regression models we calculated rate ratios (RR) with 95% confidence intervals (CI) for annual STI counts in relation to each of the exposure and sociodemographic variables for each state during the study period (see Appendix B, Tables 5-7). This was accomplished using the Proc Glimmix function in SAS version 9.4 (SAS Institute, Cary, NC, USA). The log of county population estimates for each county year was included as an offset term to account for the impact that variability in population density would have on this analysis. Finally, the sociodemographic covariates were standardized on a state-by-state basis by subtracting the mean and dividing by the standard deviation in order to create stability during model fitting.

In multivariable models, RR with 95% CI were calculated for the association between county year shale activity and cases of chlamydia, gonorrhea, and syphilis for each state, while adjusting for potential confounders. This was accomplished using Proc Glimmix with Poisson regressions in SAS. Proc Glimmix fits many different generalized linear mixed models, which are commonly used in epidemiological analyses involving longitudinal data including disease counts. Each model included a log of county population for each county year as an offset term. A county level random effect was included to establish a baseline for each county as well as to control for the potential correlation of rates across time within each county. An observation-level random effect was also included in each model to control for any excess variability in the STI case data that was unexplained by the included covariates (i.e. over dispersion). Year was also included in each model as a categorical variable to account for the unassociated rise in STI over time, with the reference year being the earliest year for which STI data was available.

Shale drilling activity was the primary independent variable of interest for all STI, and so it was included in all models. When two sociodemographic covariates were highly correlated with each other ($|r_{Spearman}| \geq 0.7$) one was removed from the model to avoid estimation troubles caused by multicollinearity. Models were kept consistent within each state for all STI and all covariates were standardized on a state-by-state basis prior to inclusion.

Colorado. Percent male was removed due to a perfect negative correlation ($r_{Spearman}$ = -1.0) with percent female. Percent white was removed due to a strong negative correlation ($r_{Spearman}$ = -0.8) with percent Hispanic or Latino. Percent Bachelor's degree was removed due to a strong negative correlation ($r_{Spearman}$ = -0.8) with percent high school. Median household income was removed due to a strong negative correlation ($r_{Spearman}$ = -0.7) with percent poverty. Percent Asian and percent American Indian or Alaska Native were removed due to the low percent of the population identifying as such. Removal of these two variables did not significantly affect modelled RR. The model for each STI in Colorado therefore included shale drilling activity (Cat_TAS_targeting_Shale), population density, percent female, percent population 15-29 years of age, percent black or African American, percent Hispanic or Latino, percent population ≥ 25 that graduated high school, percent population in poverty within the last 12 months, percent population with health insurance, and year.

North Dakota. Percent male was removed due to a perfect negative correlation ($r_{Spearman}$ = -1.0) with percent female. Percent American Indian or Alaska Native was kept in the final models despite a strong negative correlation ($r_{Spearman}$ = -0.8) with percent white, due to the high percentage of the population identifying as such. Percent Asian was removed due to the small percent of the population identifying as such. Removal of this variable did not significantly affect modelled RR. The model for each STI in North Dakota therefore included shale drilling activity (Cat_TAS_Targeting_Shale), population density, percent female, percent population 15-29 years of age, percent white, percent black or African American, percent Hispanic or Latino, percent American Indian or Alaska Native, percent population ≥ 25 that graduated high school, percent population ≥ 25 with a

Bachelor's degree, percent population in poverty within the last 12 months, percent population with health insurance, median household income, and year.

Texas. Percent male was removed due to a perfect negative correlation ($r_{Spearman} = -1.0$) with percent female. Percent Asian and percent American Indian or Alaska Native were removed due to the low percent of the population identifying as such. Removal of these two variables did not significantly affect modelled RR. The model for each STI in Texas therefore included shale drilling activity (Cat_Tas_Targeting_Shale), population density, percent female, percent population 15-29 years of age, percent white, percent black or African American, percent Hispanic or Latino, percent population ≥ 25 that graduated high school, percent population ≥ 25 with a Bachelor's degree, percent population in poverty within the last 12 months, percent population with health insurance, median household income, and year.

Results

Colorado

Drilling activity in Colorado exhibited a bimodal distribution across the study period. A marked increase in the number of active spuds targeting a shale formation was seen beginning in 2005. This increase tapered down to a low in 2009 before increasing again to a height of maximum activity in 2014. Meanwhile, averaged annual rates of chlamydia for the state increased throughout the study period. Averaged annual rates of gonorrhea remained fairly consistent from 2000-2008. A decrease in averaged annual rates of gonorrhea throughout Colorado can be seen from 2009-2013, before rates begin increasing again. Averaged annual rates of syphilis increased throughout the study period, but this increase was marginal compared to rates of chlamydia and gonorrhea. These averaged annual STI rates in Colorado follow national trends throughout the study period. (See Appendix A, Figure 8)

The study period includes 1,088 Colorado county years (64 counties x 17 years each). Of these county years, 17 (1.6%) were classified as having high shale drilling activity, 133 (12.2%) were classified as having low shale drilling activity, and 938 (86.2%) were classified as having no shale drilling activity. Population density (in ppsm) was highest for non-shale county years (156.5) followed by low (54.3) and high shale county years (30.2).

The percent of the population identifying as Hispanic or Latino was highest for high shale county years (21.7%), followed by non-shale (18.1%) and low shale county years (16.5%). Annual rates of Chlamydia (per 100,000 population) were highest for high shale county years (391.8) followed by low shale (321.0) and non-shale county years (279.3). Annual rates of Gonorrhea (per 100,000 population) were highest for non-shale county years (54.8) followed by low shale (35.9) and high shale county years (33.4). Annual rates of syphilis were highest for non-shale county years (1.8) followed by low (1.5) and high (1.5) shale county years. Among all sociodemographic and outcome variables, statistically significant differences between county year classes were observed for percent male ($p = 0.015$), percent female ($p = 0.015$), percent white ($p = 0.017$), percent high school ($p < 0.001$), and percent Bachelor's degree ($p < 0.001$) . (See Appendix B, Table 4)

Chlamydia. Relationships between shale activity and rates of chlamydia were non-monotonic and not statistically significant. In unadjusted models, the RR was elevated in the low shale activity category (RR = 1.07; 95% CI = 0.96, 1.19), but suggested an inverse association in the high shale activity category (RR = 0.86; 95% CI = 0.64, 1.15). After adjusting for sociodemographic covariates, the relationship was null for the low shale activity category (RR = 1.01; 95% CI = 0.94, 1.07), while the high shale activity category experienced an 18% (RR = 0.82; CI = 0.69, 0.99) decrease.

Gonorrhea. The relationship between shale activity and gonorrhea was null in both unadjusted and adjusted models. In unadjusted models, the RR was null for the low shale activity category (RR = 1.00; 95% CI = 0.83, 1.20), and was only slightly elevated for the high shale activity category (RR = 1.03; 95% CI = 0.64, 1.67). After adjusting for sociodemographic covariates, the relationship remained null for both the low (RR = 1.06; 95% CI = 0.91, 1.22) and high (RR = 1.01; 95% CI = 0.70, 1.45) shale activity categories.

Syphilis. The unadjusted models estimated that the low shale activity category had a 74% (RR = 1.74; 95% CI = 1.19, 2.55) increased rate of syphilis while the high shale activity category had a 406% (RR = 4.06; 95% CI = 1.61, 10.26) increased rate. However, after controlling for sociodemographic covariates: the low shale activity category had a 21% (RR = 1.21; 95% CI = 0.93, 1.57) increased rate of syphilis while the high shale activity category experienced a 33% (RR = 1.33; 95% CI = 0.65, 2.73) increase.

North Dakota

Shale drilling first began in North Dakota in 2005 and increased steadily – except for a slight decrease in 2009 – until reaching its maximum height of activity in 2014. Meanwhile, averaged annual rates of all STI increased throughout the study period, with the most pronounced increase seen for chlamydia and gonorrhea respectively. While rates of syphilis increased, the increase was marginal compared to that of chlamydia and gonorrhea. Although lower than national averages, the increased rates of chlamydia in North Dakota appear similar to the national temporal trend. Rates of gonorrhea throughout the state were also below national averages throughout the study period and began increasing at a greater rate in 2009. (Appendix A, Figure 9).

The study period includes 901 North Dakota county years (53 counties x 17 years each). Of these county years, 43 (4.8%) were classified as having high shale drilling activity, 61 (6.8%) were classified as having low shale drilling activity, and 797 (88.5%) were classified as having no shale drilling activity. Population density (in ppsm) was highest for non-shale county years (9.1), followed by high shale (5.1) and low shale (4.6) county years. The percent of the population identifying as American Indian or Alaska Native was highest for high shale county years (13.2%), followed by non-shale (6.3%) and low shale (5.5%) county years. Median household income (in USD) was highest for high shale county-years ($58,808.60), followed by low shale ($50,646.20) and non-shale county years ($41,356.20). Annual rates of chlamydia (per 100,000 population) were highest for high shale county years (45.6), followed by non-shale (45.1) and low shale county years (25.1). Annual rates of gonorrhea were highest for non-shale county years (6.2), followed by high shale (4.7) and low shale (2.6) county years. Annual rates of syphilis (0.1) were consistent for all three county year classifications. Among all sociodemographic and outcome variables, statistically significant differences between county year classes were observed for population density ($p = 0.021$), percent male ($p < 0.001$), percent female ($p < 0.001$), percent 15-29 ($p = 0.042$), percent white ($p = 0.003$), percent Asian ($p = 0.005$), percent Hispanic or Latino ($p < 0.001$), percent American Indian or Alaska Native ($p = 0.024$), median household income ($p < 0.001$), percent below poverty ($p = 0.020$), and percent health insurance ($p < 0.001$). (See Appendix B, Table 4)

Chlamydia. The relationship between shale activity and chlamydia was monotonic in both the unadjusted and adjusted models; however, the only statistically significant RR occurred with the high shale activity category using the unadjusted model. In unadjusted models the RR was elevated for both the low (RR = 1.31; 95% CI = 1.00, 1.71) and high (RR = 1.88; 95% CI = 1.38, 2.55) shale activity categories. After adjusting for sociodemographic covariates, an inverse association was observed for both the low (RR = 0.89; 95% CI = 0.72, 1.11) and high (RR = 0.96; 95% CI = 0.73, 1.27) shale activity categories.

Gonorrhea. The relationship between shale activity and gonorrhea was monotonic in both the unadjusted and adjusted models; however, the only statistically significant RR occurred with the high shale activity category using the unadjusted model. In unadjusted models, the RR was elevated for both the low (RR = 1.31; 95% CI = 0.76, 2.27) and high (RR = 2.25; 95% CI = 1.21, 4.19) shale activity categories. These effect estimates were greatly attenuated after adjusting for sociodemographic covariates: the low shale activity category had a 4% (RR = 1.04; 95% CI = 0.67, 1.62) increased rate of gonorrhea while the high shale activity category experienced a 12% (RR = 1.12; 95% CI = 0.64, 1.95) increase.

Syphilis. The relationship between shale activity and syphilis was nonmonotonic and not statistically significant in either the unadjusted or adjusted models. In unadjusted models, the RR was elevated for both the low (RR = 2.03; 95% CI = 0.52, 7.99) and high (RR = 1.62; 95% CI = 0.38, 6.94) shale activity categories. An inverse association, however, was observed after adjusting for sociodemographic covariates: the low shale activity category had a 7% (RR = 0.93; 95% CI = 0.14, 6.37) decreased rate of syphilis while the high shale activity category experienced a 37% (RR = 0.63; 95% CI = 0.09, 4.25) decrease.

Texas

Although shale drilling activity existed in Texas prior to the study period, a marked increase was seen between 2006-2008. This increase tapered down to a low in 2009 before increasing again to a height of maximum activity in 2014. Meanwhile, rates of chlamydia, gonorrhea, and syphilis increased across the study period and appear fairly consistent with national trends. (Appendix A, Figure 10)

The study period includes 4,318 Texas county years (254 counties x 17 years each). Of these county years, 141 (3.3%) were classified as having high shale drilling activity, 481 (11.1%) were classified as having low shale drilling activity, and 3,696 (86%) were classified as having no shale drilling activity. Population density (in ppsm) was highest for high shale county years (291.7), followed by low shale (99.7) and non-shale county years (87.1). The percent of the population identifying as Hispanic or Latino was highest in high shale county years (37.1%), followed by low shale (32.0%) and non-shale county years (30.1%). High shale county years also had the highest median household income (in USD) ($47,712.80), followed by low shale ($42,229.40) and non-shale county years ($40,176.80). Annual rates of chlamydia (per 100,000 population) were highest for high shale county-years (908.6), followed by low shale (419.0) and non-shale county years (371.4). High shale county years also had higher rates of gonorrhea (per 100,000 population) (280.7), followed by low shale (129.3) and non-shale county years (117.2). Rates of syphilis (per 100,000 population) were also highest in high shale county years (12.2), followed by low shale (4.4) and non-shale county years (4.3). Among all sociodemographic and outcome variables, statistically significant differences between county year classes were observed for population density ($p < 0.001$), percent male ($p < 0.001$), percent female ($p < 0.001$), percent 15-29 ($p < 0.001$), percent Asian ($p < 0.001$), percent Hispanic or Latino ($p < 0.001$), percent American Indian or Alaska Native ($p < 0.001$), percent high school graduates ($p < 0.001$), percent Bachelor's degree ($p < 0.001$), median household income ($p < 0.001$), chlamydia rate ($p < 0.001$), gonorrhea rate ($p = 0.004$), and syphilis rate ($p = 0.006$). (See Appendix B, Table 4)

Chlamydia. The relationship between shale activity and chlamydia was monotonic in both the unadjusted and adjusted models. In unadjusted models, the RR was elevated for both the low (RR = 1.14; 95% CI = 1.09, 1.19) and high (RR = 1.36; 95% CI = 1.26, 1.47) shale activity categories. After adjusting for sociodemographic covariates the low shale activity category had a 2% (RR = 1.02; 95% CI = 0.99, 1.06) increased rate of chlamydia while the high shale activity category experienced a 10% (RR = 1.10; 95% CI = 1.04, 1.17) increase.

Gonorrhea. The relationship between shale activity and gonorrhea was monotonic in both the unadjusted and adjusted models. In unadjusted models, the RR was elevated for both the low (RR = 1.02; 95% CI = 0.96, 1.08) and high (RR = 1.13; 95% CI = 1.02, 1.26) shale activity categories. After adjusting for sociodemographic covariates the low shale activity category had a 2% (RR = 1.02; 95% CI = 0.96, 1.08) increased rate of gonorrhea while the high shale activity category experienced a 15% (RR = 1.15; 95% CI = 1.04, 1.28) increase.

Syphilis. The relationship between shale activity and gonorrhea was monotonic in unadjusted models and nonmonotonic in the adjusted models. In unadjusted models, the RR was elevated for both the low (RR = 1.23; 95% CI = 0.91, 1.67) and high (RR = 1.33; 95% CI = 0.79, 2.25) shale activity categories. After adjusting for sociodemographic covariates the low shale activity category had a 6% (RR = 0.94) decreased rate of syphilis while the high shale activity category experienced a 12% (RR = 0.88) decrease.

Discussion

This analysis was motivated by previous ecologic studies that observed positive associations between annually reported STI and shale activity in communities overlying the Marcellus Shale. The purpose of this analysis, therefore, was to assess whether this previously observed relationship is a unique place-based phenomenon or a generalizable result of UOGD. Compared to county years with no shale activity, Texas county years with high shale activity (≥ 50 shale spuds) had 10% (RR = 1.10; 95% CI = 1.04, 1.17) increased rates of chlamydia and 15% (RR = 1.15; 95% CI = 1.04, 1.28) increased rates of gonorrhea. No association was observed for syphilis in Texas. An association between shale activity and annually reported rates of chlamydia, gonorrhea, and syphilis was not observed in Colorado and North Dakota.

The findings in Texas support previous research that found an increased rate of chlamydia and gonorrhea in counties with high shale activity (Deziel et al., 2018; Komarek & Cseh, 2017). Incongruous findings between Texas, Colorado, and North Dakota may be due the higher number of Texas county year observations for all shale activity categories, which may have increased the ability to detect an association in the state.

The incongruous findings between this and previous studies may suggest the presence of interstate or regional differences in the impact of shale activity. For instance, while Deziel et al. (2018) observed a low number of new well permits in Ohio until a spike in 2010, the annual frequency of active spuds targeting a shale formation in all three states in this study was much more variable. What was common to each state, however, was a decline in new spuds in 2009, which may have been due to the effects of the 2008 U.S. financial crisis. Furthermore, an association may exist between annual STI rates and wells targeting a non-shale formation, which were not included in this analysis. There also may be regional differences in the proximity of oil and gas wells in relation to population dense towns and cities. While population density was controlled for in each model, a more robust assessment of this relationship may be needed in order to adequately quantify and control for the intensity of interaction between drilling workers and community members.

One of the primary strengths of this analysis was the inclusion of multiple STI outcomes across multiple states and regions. This analysis also relied on multiple national databases, which protected the collection of sociodemographic, STI, and drilling information from variations in state-based reporting practices. Relying on the spud date as the indicator for shale activity is also an improved exposure metric that likely protects against exposure misclassification by better estimating both drilling activity and the time at which a large migratory workforce would be present.

There are also limitations to this study, which may have limited our ability to detect an association between shale activity and annual rates of STI. This is an ecologic analysis, and as such it is only possible to assess the presence of population-level associations. This means that the lack of a positive association in this study does not conclusively rule out an association at the individual level. Although we adjusted for a number of factors, we did not directly adjust for spatial correlation. Other limitations include the reliance on a passive STI surveillance system. Reported incidence of STI are likely underestimates and may be attributed to counties other than their true source of origin. Furthermore, our statistical power was limited in various models due to the exposure classification chosen, which limited the number of county years in all shale activity categories, especially for the high shale activity category.

This analysis was meant to observe if the association between shale activity and STI in counties overlying the Marcellus Shale could be reproduced in a different geographic region of the United States. Further studies aimed at understanding the interactions between community members and drilling workers are still needed to shed light on the strength of this association in this and other geographic areas.

Conclusions

The first multi-state, multi-region analysis of shale activity and annual STI rates confirmed previously observed associations in Texas counties, for the association between shale drilling activity and annual rates of chlamydia and gonorrhea. While elevated rates in the unadjusted models in Colorado and North Dakota suggested a positive association, these effect estimates did not remain after rigorous adjustments for sociodemographic covariates, secular trends, county-level random effects, and observation-level random effects.

References

Centers for Disease Control and Prevention. (n.d.). Atlas Plus Technical Notes for STD Surveillance Data. Available from: https://gis.cdc.gov/grasp/nchhstpatlas/main.html.

Centers for Disease Control and Prevention. (2018a). National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP) Atlas Plus tool. Available from: https://gis.cdc.gov/grasp/nchhstpatlas/main.html.

Centers for Disease Control and Prevention. (2018b). Sexually Transmitted Disease Surveillance 2017: Special Focus Profiles. Available from: https://www.cdc.gov/std/stats17/minorities.htm#ref1.

Cook, T., & Perrin, J. (2016). Today In Energy: Hydraulic Fracturing Accounts for About Half of Current U.S. Crude Oil Production. Retrieved from: https://www.eia.gov/todayinenergy/detail.php?id=25372.

Deziel, N.C., Humeau, Z., Elliott, E.G., et al. (2018). Shale gas activity and increased rates of sexually transmitted infections in Ohio, 2000-2016. *PLoS One, 13*(3). Retrieved from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0194203.

Drillinginfo. (2019). DI Application. Available from: https://info.drillinginfo.com/.

Dunn, C., & Hess, T. (2018). Today In Energy: The United States is Now The Largest Global Crude Oil Producer. Retrieved from: https://www.eia.gov/todayinenergy/detail.php?id=37053# .

EIA. (2016). Summary Maps: Shale Gas and Oil Plays, Lower 48 States. Available from: https://www.eia.gov/maps/maps.htm.

EIA. (2018). Today In Energy: United States Remains The World's Top Producer of Petroleum and Natural Gas Hydrocarbons. Retrieved from: https://www.eia.gov/todayinenergy/detail.php?id=36292 .

Elliott, E.G., Ma, X., Leaderer, B.P., et al. (2018). A community-based evaluation of proximity to unconventional oil and gas wells, drinking water contaminants, and health symptoms in Ohio. *Environmental Research, 167*(), 550-557. Retrieved from: https://www.sciencedirect.com/science/article/pii/S0013935118304596.

Hays, J., McCawley, M., & Shonkoff, S.B.C. (2017). Public health implications of environmental noise associated with unconventional oil and gas development. *Science of the Total Environment, 580*(), 448-456. Retrieved from: https://www.sciencedirect.com/science/article/pii/S0048969716325724.

Hirsch, J.K., Smalley, K.B., Selby-Nelson, E.M., et al. (2018). Psychosocial impact of fracking: A review of the literature on the mental health consequences of hydraulic fracturing. *International Journal of Mental Health Addiction, 16*(), 1-15. Retrieved from: https://link.springer.com/article/10.1007/s11469-017-9792-5.

Komarek, T. & Cseh, A. (2017). Fracking and public health: Evidence from gonorrhea incidence in the Marcellus Shale region. *Journal of Public Health Policy, 38*(4). Retrieved from: https://link.springer.com/article/10.1057%2Fs41271-017-0089-5.

Litovitz, A., Curtright, A., Abramzon, S., et al. (2013). Estimation of regional air-quality damages from Marcellus Shale natural gas extraction in Pennsylvania. *Environmental Research Letters, 8*(1). Retrieved from:

https://iopscience.iop.org/article/10.1088/1748-9326/8/1/014017/pdf.

Moniz, E. (2014). Energy Secretary Ernest Moniz's Statement to the Senate Committee on Appropriations on Driving Innovation Through Federal Investments. Retrieved from: https://www.energy.gov/articles/energy-secretary-ernest-monizs-statement-senate-committee-appropriations-driving-innovation.

Morrone, M., Chadwick, A.E., & Kruse, N. (2015). A community divided: Hydraulic fracturing in rural Appalachia. *Journal of Appalachian Studies, 21*(2), 207-228. Retrieved from: https://www.jstor.org/stable/10.5406/jappastud.21.2.0207#metadata_info_tab_contents.

Perrin, J., & Cook, T. (2016). Today In Energy: Hydraulically Fractured Wells Provide Two-Thirds of U.S. Natural Gas Production. Retrieved from: https://www.eia.gov/todayinenergy/detail.php?id=26112 .

Udoh, I.A., Stammen, R.M., Mantell, J.E. (2008). Corruption and oil exploration: expert agreement about the prevention of HIV/AIDS in the Niger Delta of Nigeria. *Health Education Research, 34(4)*, 670-681. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2733797/.

United States Census Bureau American Fact Finder. (2017). Available from: https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml.

Westwood, E., Orenstein, M. (2016). Does resource development increase community sexually transmitted infections? An environmental scan. *The Extractive Industries and Society, 3(1),* 240-248. Retrieved from: https://www.sciencedirect.com/science/article/pii/S2214790X15001422.

Appendix A

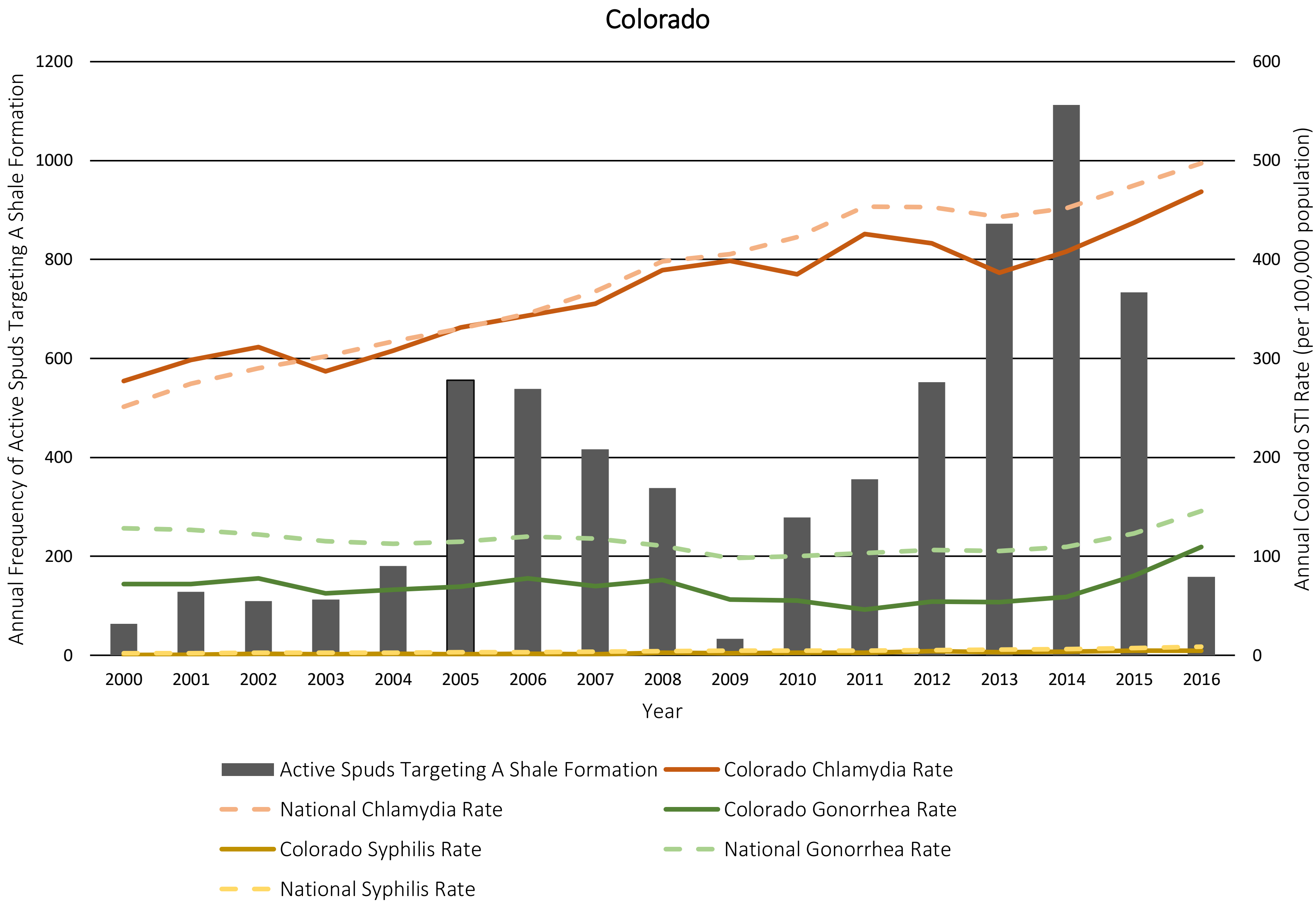

Figure 1: Annual Colorado STI rates and annual frequency of active shale spuds targeting a shale formation 2000-2016.

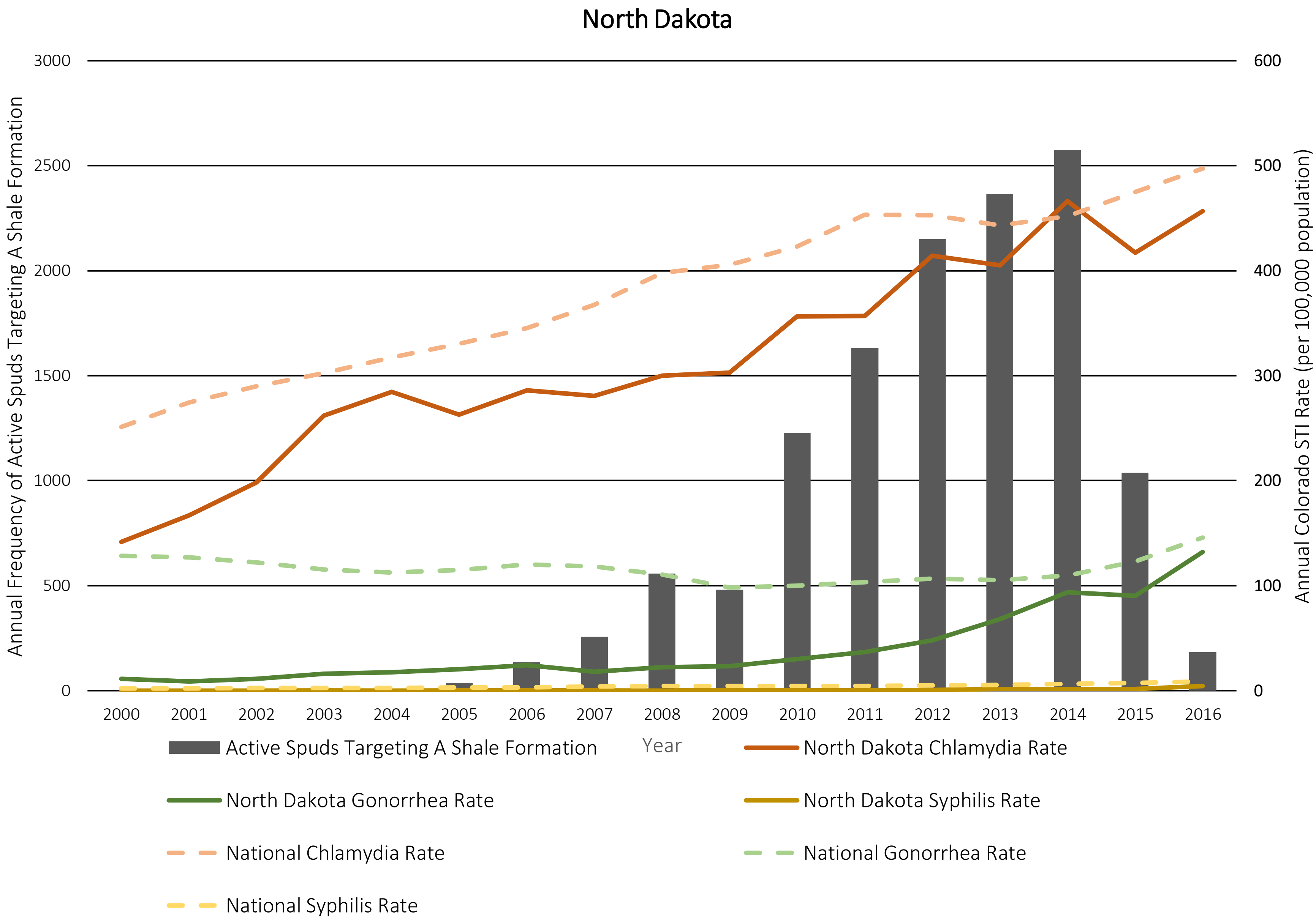

Figure 2: Annual North Dakota STI rates and annual frequency of active shale spuds targeting a shale formation 2000-2016.

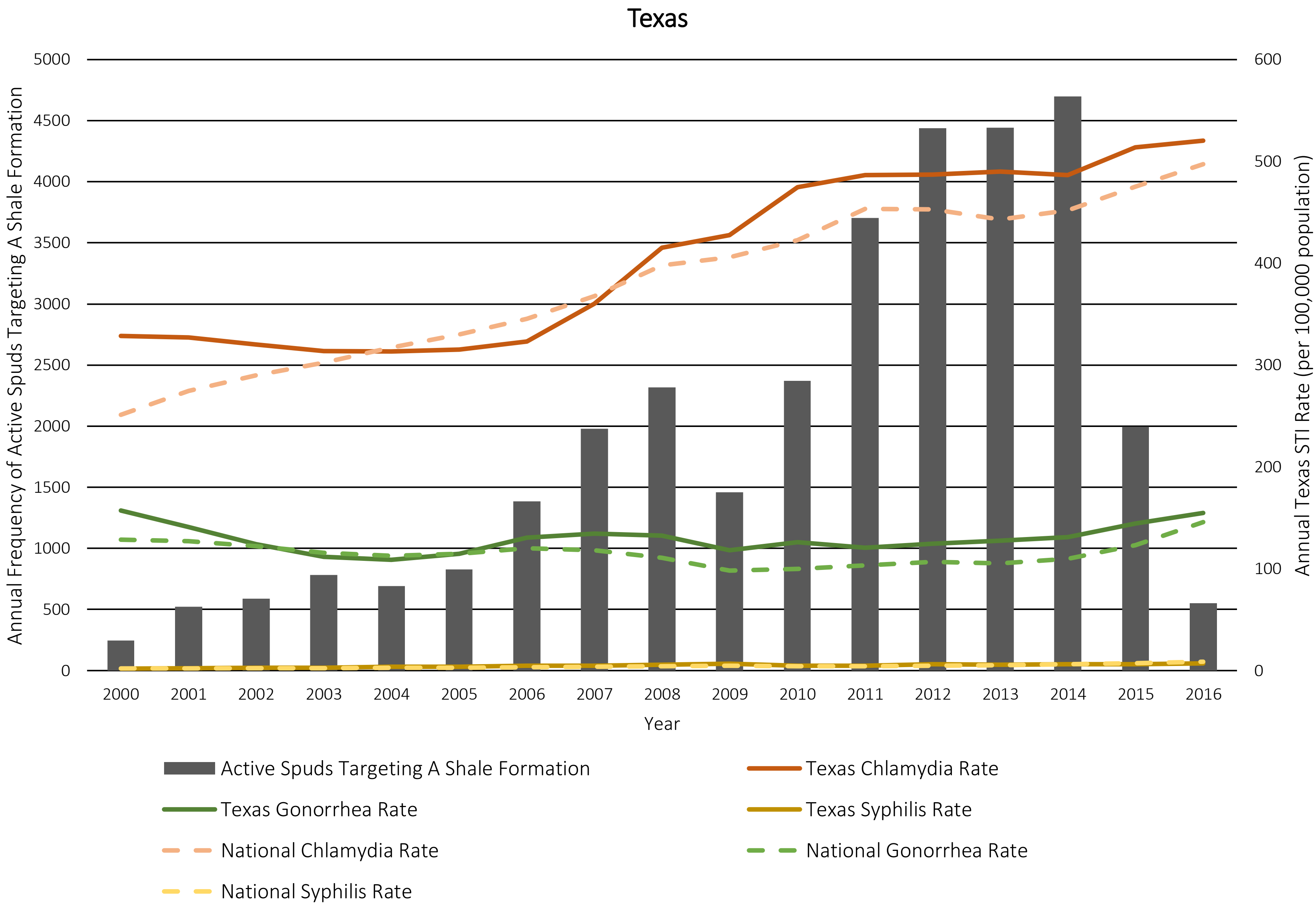

Figure 3: Annual Texas STI rates and annual frequency of active shale spuds targeting a shale formation 2000-2016.

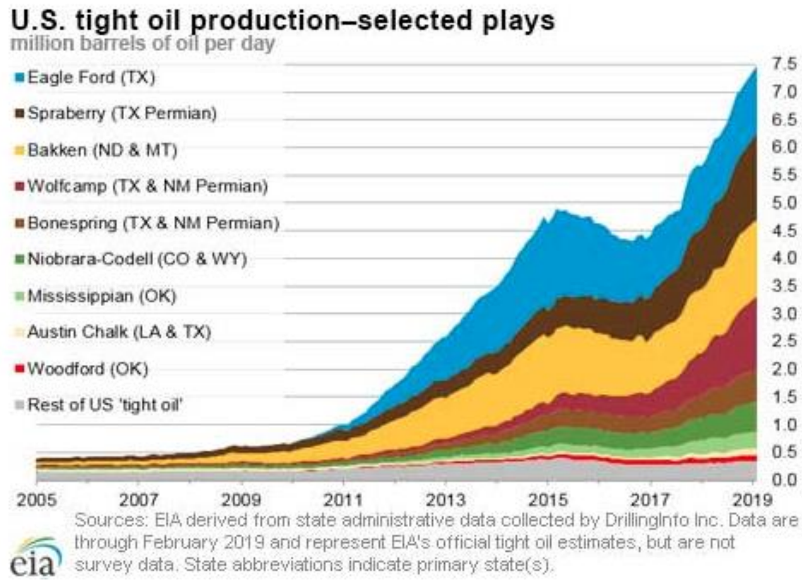

Figure 4 U.S. tight oil production – selected plays. Retrieved from: https://www.eia.gov/petroleum/data.php#crude.

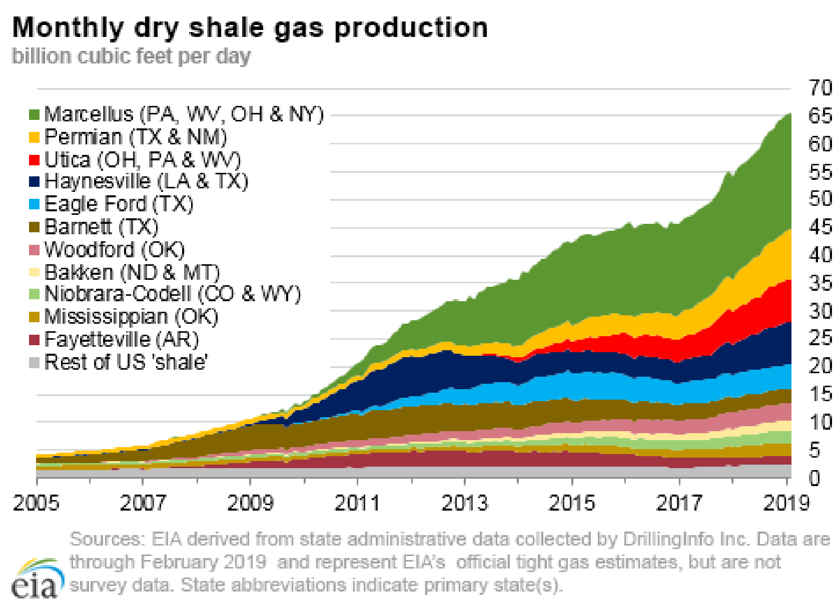

Figure 5 Monthly dry shale gas production. Retrieved from: https://www.eia.gov/naturalgas/data.php.

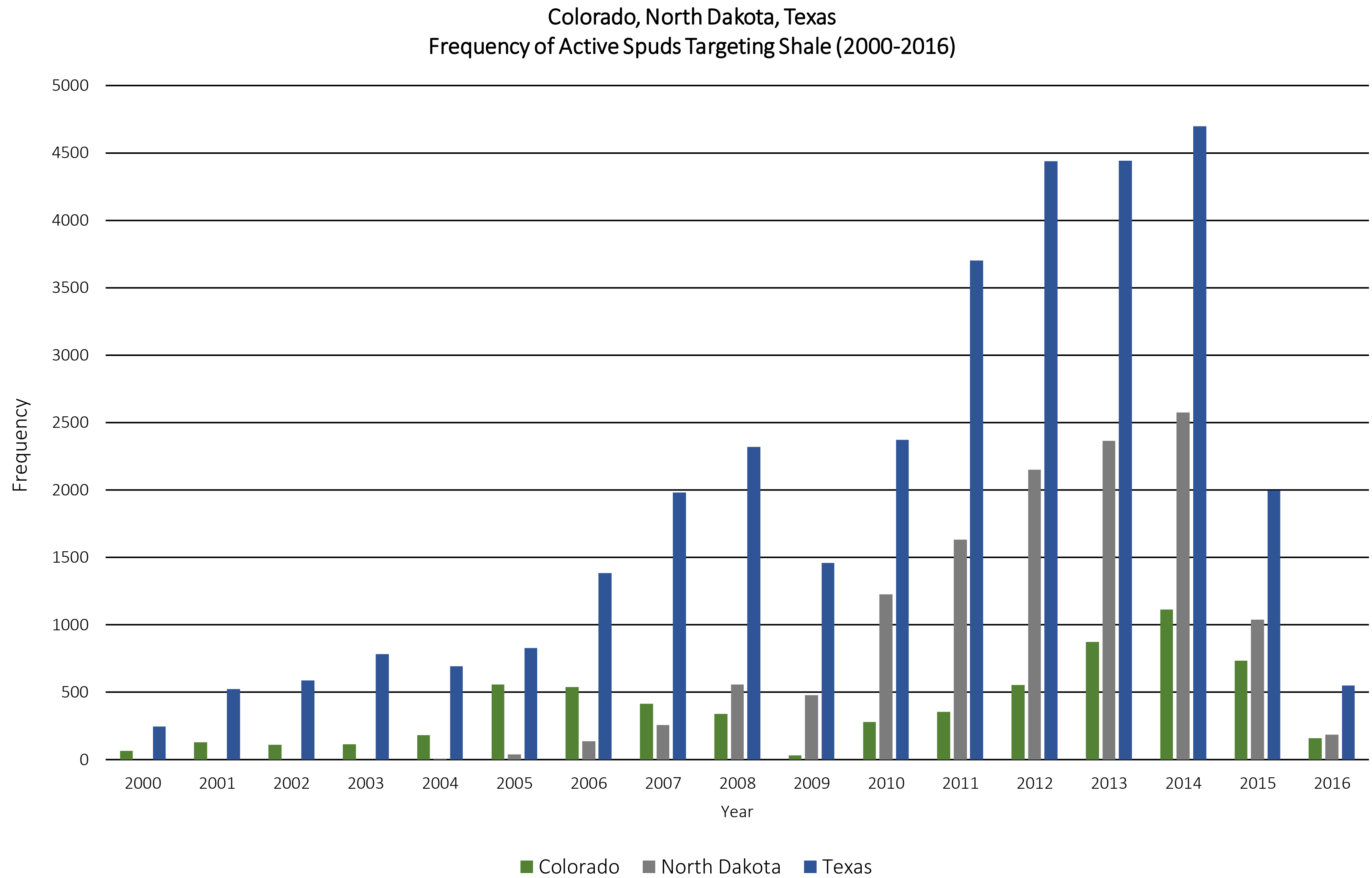

Figure 6: Annual frequency of active shale spuds in Colorado, North Dakota, and Texas (2000-2016)

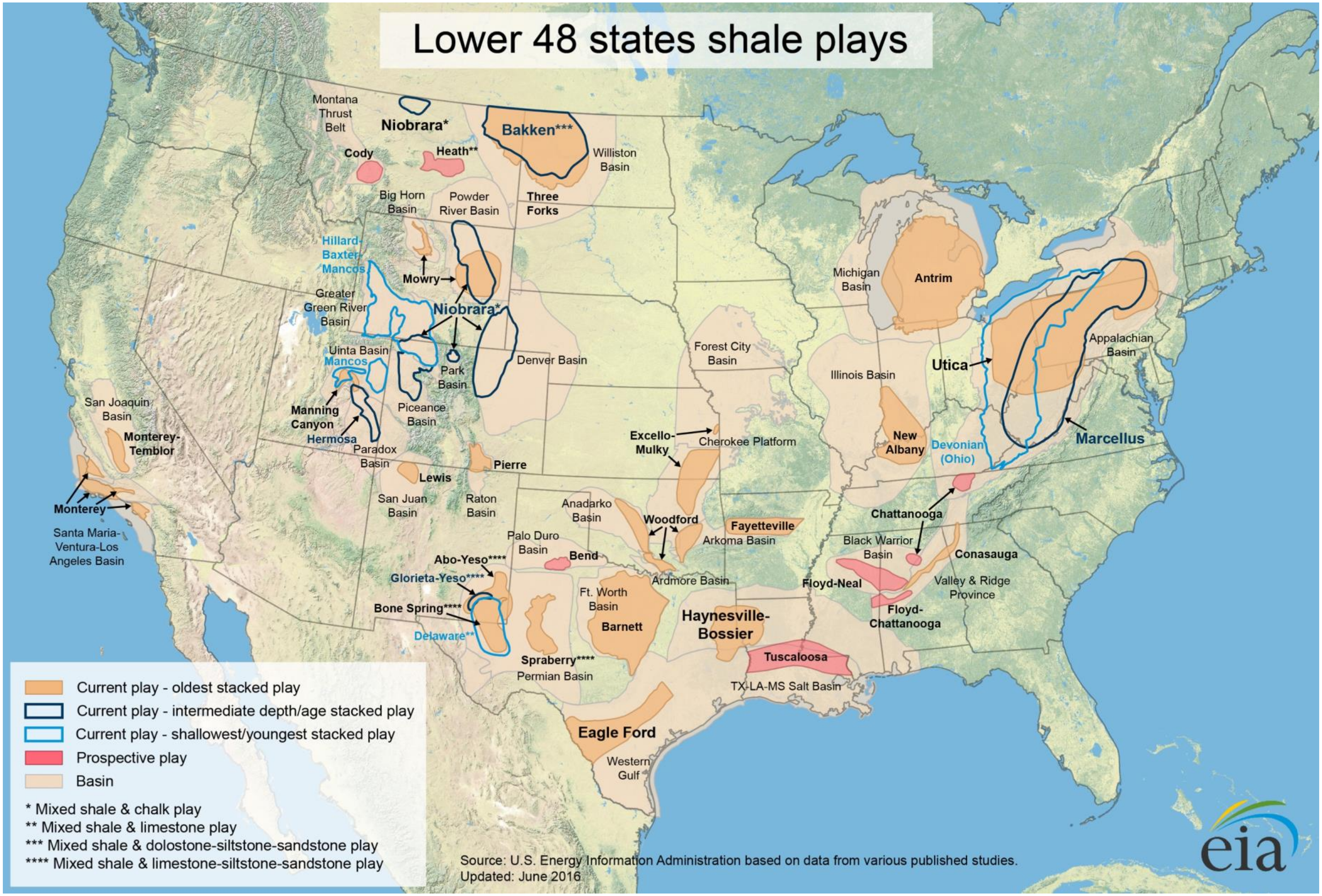

Figure 7: 2016 EIA map of shale gas and oil plays in the lower 48 United States.

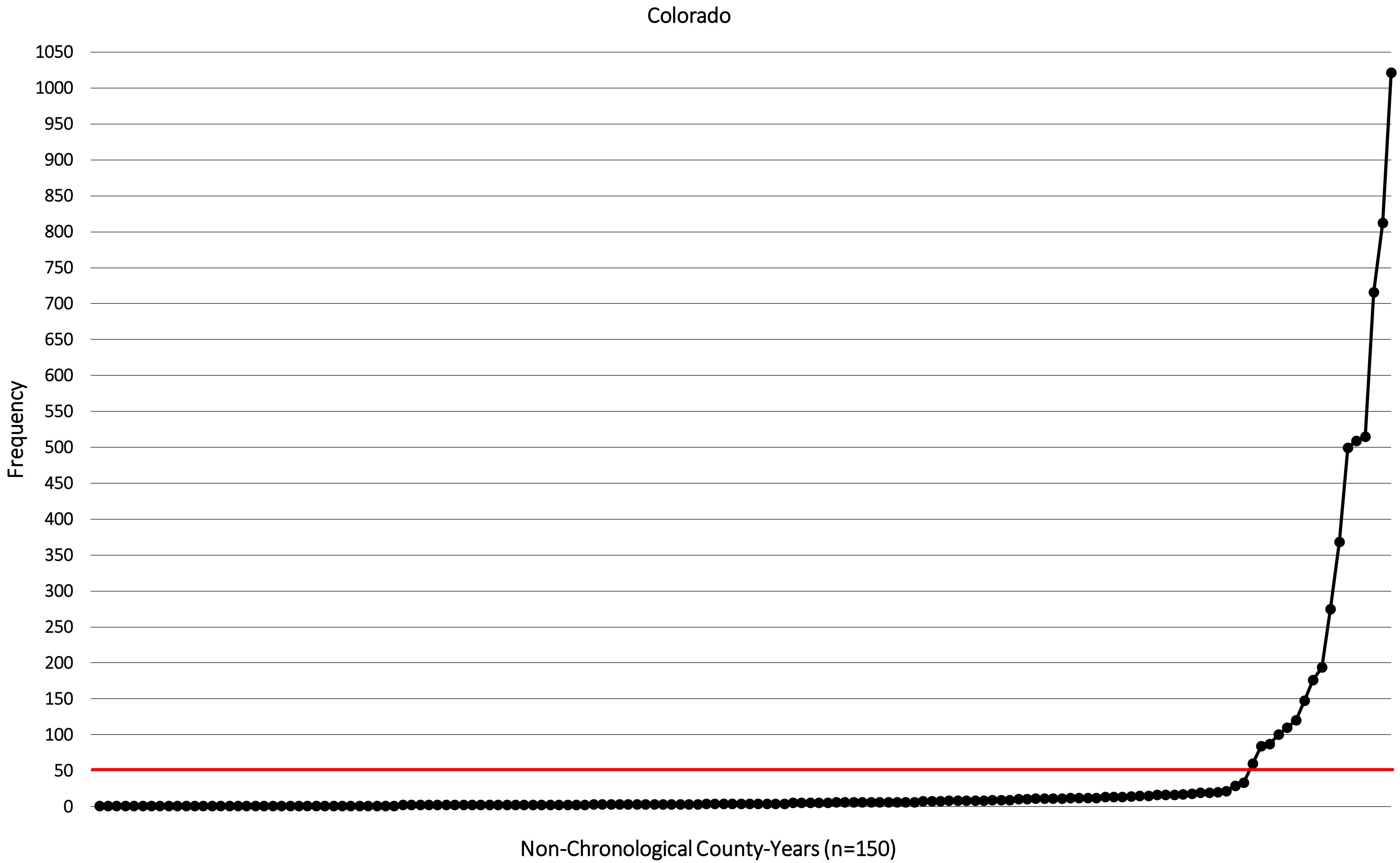

Figure 8: Frequency of active shale spuds per county year in Colorado (2000-2016). County years are shown in non-chronological order to highlight the distribution of shale drilling activity. There are a total of 1,088 county years for Colorado (2000-2016). County years shown above (n=150) include those with at least 1 active shale spud (county years with a shale spud frequency of zero have been excluded). Minimum (1), maximum (1,021), standard deviation (143.48), median (4.00).

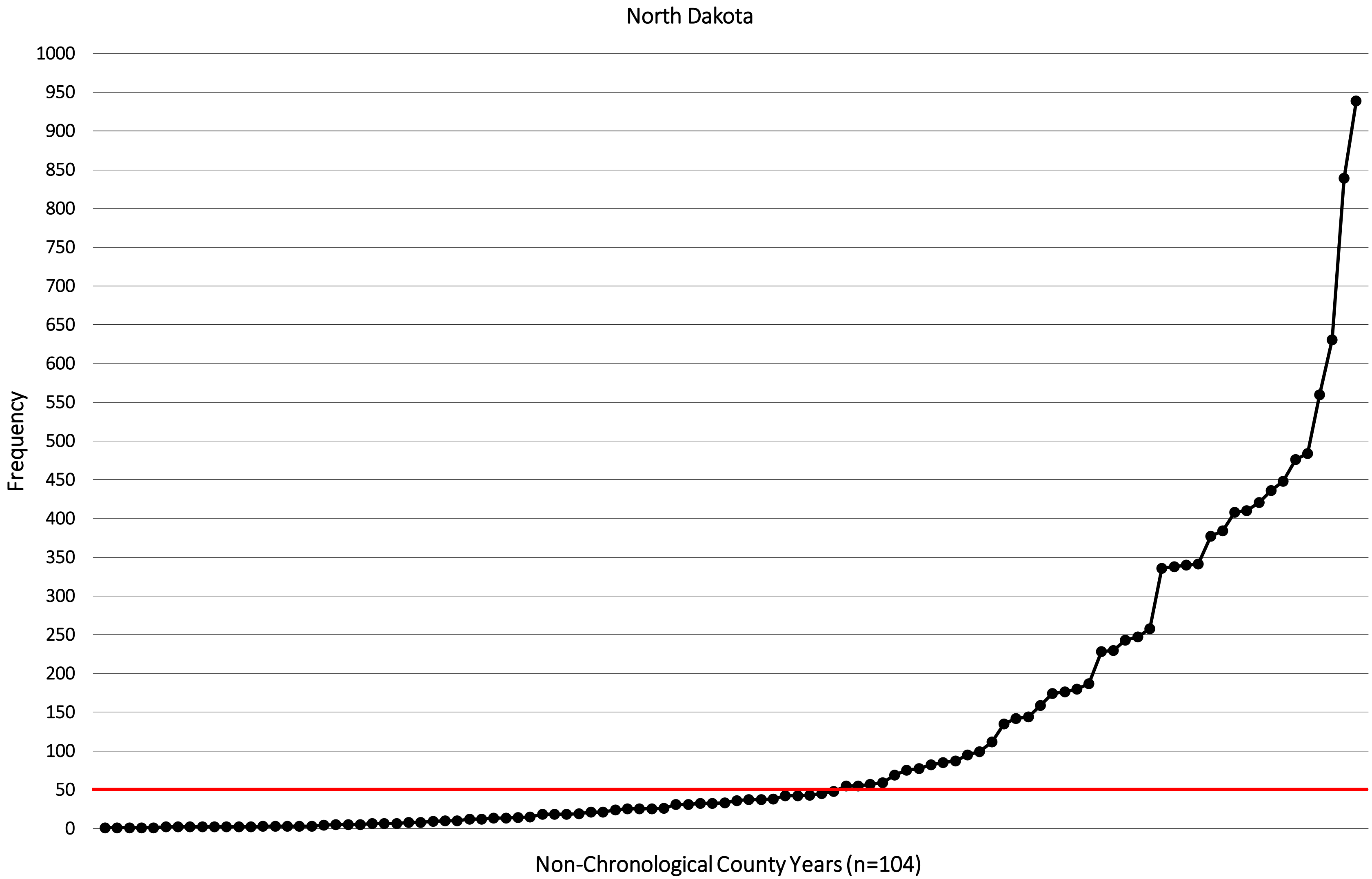

Figure 9: Frequency of active shale spuds per county year in North Dakota (2000-2016). County years are shown in non-chronological order to highlight the distribution of shale activity. There are a total of 901 county years for North Dakota (2000-2016). County years shown above (n=104) include those with at least 1 active shale spud (county years with a value of zero have been excluded). Minimum (1), maximum (939), standard deviation (183.98), median (34.50).

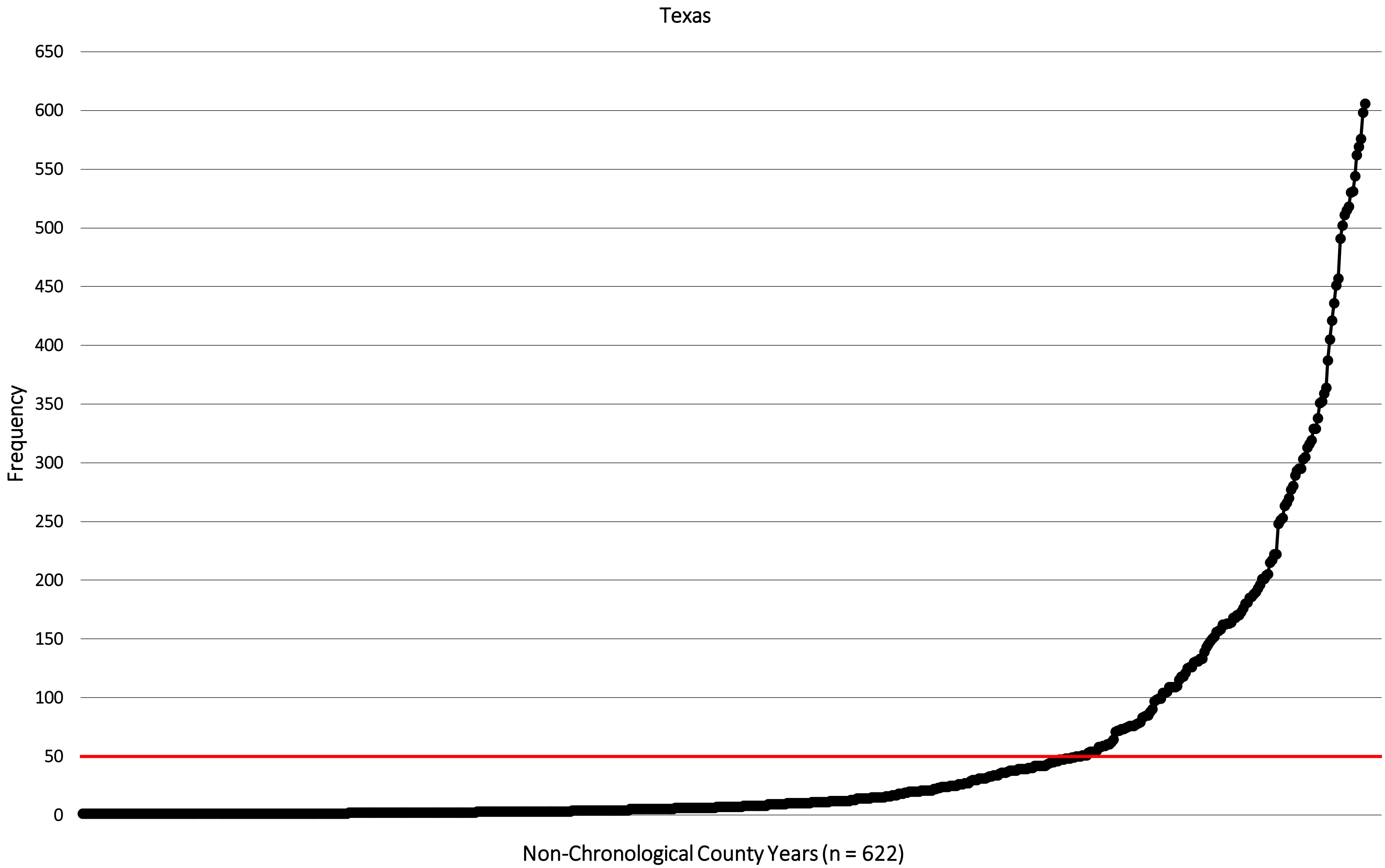

Figure 10: Frequency of active shale spuds per county year in Texas (2000-2016). County years are shown in non-chronological order to highlight the distribution of shale activity. There are a total of 4,318 county years for Texas (2000-2016). County years shown above (n=622) include those with at least 1 active shale spud (county years with a value of zero have been excluded). Minimum (1), maximum (606), standard deviation (107.40), median (7.00).

Appendix B

Table 1: Census dataset for covariates (2000-2016). DEC refers to the decennial census. ACS refers to the American Community Census. *Excluding % Insured and County Area.

	Covariates*	% Insured
2000-2004	2000 DEC Census	2012 ACS (5-yr estimates)
2005-2009	2009 ACS Survey (5-yr estimates)	2012 ACS (5-yr estimates)
2010	2010 ACS (5-yr estimates)	2012 ACS (5-yr estimates)
2011	2011 ACS (5-yr estimates)	2012 ACS (5-yr estimates)
2012	2012 ACS (5-yr estimates)	
2013	2013 ACS (5-yr estimates)	
2014	2014 ACS (5-yr estimates)	
2015	2015 ACS (5-yr estimates)	
2016	2016 ACS (5-yr estimates)	

Table 2: Census covariate sources (2000-2016). DEC refers to the decennial census. ACS refers to the American Community Census.

	2000-2004	2005-2016
% Male	DEC: P012	ACS: S0101 (5-yr estimates)
% Female	DEC: P012	ACS: S0101 (5-yr estimates)
% 15-29	DEC: P012	ACS: S0101 (5-yr estimates)
% White	DEC: DP1 SF1	ACS: B02001 (5-yr estimates)
% Black or Afr. Am. Amrican	DEC: DP1 SF1	ACS: B02001 (5-yr estimates)
% Am. In. & AK Nat.	DEC: DP1 SF1	ACS: B02001 (5-yr estimates)
% Asian	DEC: DP1 SF1	ACS: B02001 (5-yr estimates)
% Hispanic or Latino	DEC: DP1 SF1	ACS: B03003 (5-yr estimates)
% HS Graduate	DEC: DP2 SF3	ACS: S1501 (5-yr estimates)
% Bachelor's Degree	DEC: DP2 SF3	ACS: S1501 (5-yr estimates)
Med. Household Income	DEC: DP2 SF3	ACS: S1901 (5-yr estimates)
% Indiv. < Pov. Level	DEC: DP2 SF3	ACS: B17001 (5-yr estimates)
% Insured	ACS: B27001 (5-yr estimates)	
County Area	GCT-PH1 2010 CSF 1	

Table 3: Shale target formation frequency by state 2000-2016.

	Target Formation	Frequency
Colorado	Niobrara Shale	6311
	Baxter-Mancos Shale	165
	Pierre Shale	67
North Dakota	Bakken Shale	8278
	Three Forks Shale	4364
Texas	Eagle Ford Shale	15,961
	Barnett Shale	14,517
	Haynesville Shale	943
	Bone Spring Shale	546
	Delaware Shale	480
	Spraberry Shale	351
	Bossier Shale	139
	Shale	19
	Bend Shale	15
	Woodford Shale	15
	Black Shale	3
	Sandy Shale	2
	Anhydrite Shale	1
	CV Shale	1
	Jackson Shale	1
	Pregnant Shale	1
	Smithwick Shale	1
	Valera Shale	1
	Wolfcamp Shale	1
	Wolfcamp Shale	1
	Wolfcamp Shale	1

Table 4: Distribution of exposure, sociodemographic, and outcome variables among county years with no shale spuds, < 50 shale spuds, and ≥ 50 shale spuds in Colorado, North Dakota, and Texas (2000-2016).

	Colorado County Years				North Dakota County Years				Texas County Years			
	No Shale Spuds (n = 938)	< 50 Shale Spuds (n = 133)	≥ 50 Shale Spuds (n = 17)		No Shale Spuds (n = 797)	< 50 Shale Spuds (n = 61)	≥ 50 Shale Spuds (n = 43)		No Shale Spuds (n = 3,696)	< 50 Shale Spuds (n = 481)	≥ 50 Shale Spuds (n = 141)	
Exposure	Mean (SD)	Mean (SD)	Mean (SD)	P-Value	Mean (SD)	Mean (SD)	Mean (SD)	P-Value	Mean (SD)	Mean (SD)	Mean (SD)	P-Value
Active Shale Spuds	0	5.6 (5.9)	340.8 (292.6)	--	0	15.8 (14.2)	271.6 (208.4)	--	0	9.3 (12.0)	202.4 (147.0)	--
Sociodemographic Factors												
Population Density (ppsm)	156.5 (567.9)	54.3 (142.2)	30.2 (32.1)	0.079	9.1 (15.6)	4.6 (6.1)	5.1 (5.2)	0.021	87.1 (251.9)	99.7 (373.3)	291.7 (558.4)	< 0.001
% Male	51.8 (3.9)	51.2 (2.6)	49.5 (0.6)	0.015	50.4 (1.4)	50.9 (1.9)	52.0 (1.7)	< 0.001	50.3 (3.0)	51.4 (3.9)	51.1 (4.0)	< 0.001
% Female	48.2 (3.9)	48.8 (2.6)	50.5 (0.6)	0.015	49.6 (1.4)	49.1 (1.9)	48.0 (1.7)	< 0.001	49.7 (3.0)	48.6 (3.9)	48.9 (4.0)	< 0.001
% 15-29	18.8 (4.9)	19.7 (4.0)	19.0 (2.8)	0.088	17.0 (4.9)	16.3 (4.0)	18.6 (3.7)	0.042	19.5 (4.1)	20.3 (4.5)	20.8 (3.7)	< 0.001
% White	88.8 (7.4)	90.4 (5.7)	91.7 (3.5)	0.017	91.6 (17.4)	92.5 (8.7)	82.7 (11.6)	0.003	82.5 (9.0)	82.3 (8.8)	82.1 (11.0)	0.808
% Black	1.6 (2.5)	1.6 (2.4)	2.6 (3.9)	0.225	0.8 (1.6)	0.9 (1.6)	0.9 (1.8)	0.843	6.7 (6.9)	6.8 (6.9)	5.3 (5.3)	0.050
% Asian	1.1 (1.3)	0.9 (1.3)	0.6 (0.6)	0.085	0.6 (1.1)	1.1 (1.6)	0.8 (1.1)	0.005	1.1 (1.9)	0.8 (1.6)	1.7 (2.3)	< 0.001
% Hispanic	18.1 (15.1)	16.5 (9.9)	21.7 (6.4)	0.268	1.4 (1.6)	1.6 (1.4)	2.6 (1.4)	< 0.001	30.1 (22.4)	32.0 (24.4)	37.1 (29.5)	< 0.001
% Native	1.5 (2.0)	1.3 (1.3)	0.7 (0.3)	0.106	6.3 (17.2)	5.5 (7.7)	13.2 (10.8)	0.024	0.8 (1.2)	1.4 (3.0)	1.3 (2.4)	< 0.001
% High School	27.4 (7.1)	30.6 (5.8)	31.1 (3.7)	< 0.001	31.7 (4.0)	32.9 (3.6)	31.6 (2.9)	0.063	31.6 (5.6)	33.7 (6.1)	30.7 (7.5)	< 0.001
% Bachelor's Degree	19.1 (8.9)	16.0 (6.6)	15.0 (2.7)	< 0.001	14.7 (3.7)	15.3 (2.2)	15.5 (2.2)	0.172	11.8 (4.7)	10.6 (4.3)	12.8 (6.3)	< 0.001
Median Household Income (USD)	47,064.4 (14,799.8)	47,161.0 (10,181.7)	46,249.6 (10,861.8)	0.970	41,356.2 (10,502.8)	50,646.2 (10,493.7)	58,808.6 (11,017.4)	< 0.001	40,176.8 (11,010.0)	42,229.4 (10,798.1)	47,712.8 (11,617.3)	< 0.001
% Below Poverty	13.2 (6.1)	12.4 (3.0)	12.2 (1.8)	0.324	12.8 (7.0)	10.6 (2.5)	11.3 (3.1)	0.020	17.5 (6.6)	17.1 (6.9)	16.4 (7.6)	0.083
% Health Insurance	83.0 (4.8)	82.6 (4.6)	85.0 (1.4)	0.138	90.1 (5.8)	87.9 (3.7)	86.1 (3.9)	< 0.001	78.2 (4.9)	78.2 (4.9)	77.8 (5.4)	0.655
Outcomes												
Chlamydia Rate (per 100,000 pop)	279.3 (868.4)	321.0 (667.3)	391.8 (497.3)	0.762	45.1 (109.1)	25.1 (45.9)	45.6 (52.7)	0.351	371.4 (1,614.2)	419.0 (1,935.0)	908.6 (2,002.4)	< 0.001
Gonorrhea Rate (per 100,000 pop)	54.8 (208.9)	35.9 (91.3)	33.4 (47.0)	0.541	6.2 (19.9)	2.6 (5.7)	4.7 (6.5)	0.328	117.2 (558.5)	129.3 (658.6)	280.7 (702.4)	0.004
Syphilis Rate (per 100,000 pop)	1.8 (9.9)	1.5 (4.8)	1.5 (2.2)	0.951	0.1 (0.8)	0.1 (0.4)	0.1 (0.3)	0.884	4.3 (28.4)	4.4 (27.1)	12.2 (36.4)	0.006

Table 5: Rate ratios for annual STI counts in relation to drilling and sociodemographic characteristics in Colorado (2000-2016).

	Colorado					
	Chlamydia		Gonorrhea		Syphilis	
Model Parameters	RR* (95% CI)	P-Value	RR* (95% CI)	P-Value	RR* (95% CI)	P-Value
Active Shale Spuds (ref=0)	1.00	--	1.00	--	1.00	--
Low (< 50)	1.22 (1.20, 1.23)	< 0.001	0.90 (0.86, 0.93)	< 0.001	1.63 (1.36, 1.95)	< 0.001
High (≥ 50)	1.54 (1.48, 1.60)	< 0.0010	0.83 (0.74, 0.94)	0.002	4.37 (.21, 8.64)	< 0.001
Population Density (ppsm)	1.00 (1.00, 1.00)	< 0.001	1.00 (1.00, 1.00)	< 0.001	1.00 (1.00, 1.00)	< 0.001
% Male	0.56 (0.54, 0.59)	< 0.001	1.72 (1.55, 1.91)	< 0.001	0.04 (0.02, 0.06)	< 0.001
% Female	1.77 (1.70, 1.85)	< 0.001	0.58 (0.52, 0.65)	< 0.001	27.61 (15.65, 48.70)	< 0.001
% 15-29	0.65 (0.63, 0.66)	< 0.001	1.24 (1.18, 1.31)	< 0.001	0.23 (0.17, 0.30)	< 0.001
% White	1.04 (1.03, .05)	< 0.001	0.85 (0.83, 0.87)	< 0.001	1.56 (1.41, 1.74)	< 0.001
% Black	1.36 (1.34, 1.38)	< 0.001	1.30 (1.26, 1.34)	< 0.001	1.13 (0.97, .30)	0.106
% Asian	1.46 (1.44, 1.47)	< 0.001	1.01 (0.98, 1.04)	0.435	4.14 (3.57, 4.81)	< 0.001
% Hispanic	2.02 (1.97, 2.08)	< 0.001	1.22 (1.15, 1.29)	< 0.001	3.74 (2.80, 5.00)	< 0.001
% Native	0.95 (0.94, 0.97)	< 0.001	1.17 (1.11, 1.23)	< 0.001	0.77 (0.61, 0.97)	0.030
% High School	0.75 (0.73, 0.77)	< 0.001	0.88 (0.83, 0.93)	< 0.001	0.18 (0.14, 0.23)	< 0.001
% Bachelor's Degree	1.74 (1.70, 1.78)	< 0.001	0.89 (0.85, 0.94)	< 0.001	7.16 (5.65, 9.06)	< 0.001
Median Household Income (USD)	1.44 (1.42, 1.45)	< 0.001	1.07 (1.04, 1.10)	< 0.001	3.75 (3.29, 4.28)	< 0.001
% Below Poverty	1.46 (1.44, 1.48)	< 0.001	0.99 (0.88, 0.92)	< 0.001	3.52 (3.01, 4.12)	< 0.001
% Health Insurance	1.28 (1.27, 1.29)	< 0.001	1.44 (1.40, .47)	< 0.001	2.44 (2.19, 2.71)	< 0.001
Year						
2000**	--	--	1.00 (ref)	--	1.00 (ref)	--
2001	1.00 (ref)	--	1.00 (0.96, 1.05)	0.878	2.05 (1.00, 4.21)	0.050
2002	1.06 (1.03, .08)	< 0.001	1.10 (1.05, 1.15)	< 0.001	5.68 (2.99, 10.78)	< 0.001
2003	0.98 (0.95, 1.00)	0.051	0.89 (0.84, 0.93)	< 0.001	3.36 (1.72, 6.58)	< 0.001
2004	1.06 (1.03, 1.08)	< 0.001	0.94 (0.90, 0.99)	0.021	5.46 (2.87, 10.37)	< 0.001
2005	1.12 (1.09, 1.15)	< 0.001	0.98 (0.94, 1.03)	0.541	3.92 (2.03, 7.59)	< 0.001
2006	1.12 (1.09, 1.14)	< 0.001	1.10 (1.05, 1.15)	< 0.001	5.84 (3.08, 11.04)	< 0.001
2007	1.22 (1.20, 1.25)	< 0.001	0.99 (0.95, 1.04)	0.776	4.72 (2.47, 9.02)	< 0.001
2008	1.34 (1.31, 1.37)	< 0.001	1.08 (1.03, 1.13)	0.002	10.31 (5.56, 19.11)	< 0.001
2009	1.39 (1.36, 1.42)	< 0.001	0.80 (0.76, 0.84)	< 0.001	8.47 (4.55, 15.77)	< 0.001
2010	1.34 (1.31, 1.38)	< 0.001	0.79 (0.75, 0.83)	< 0.001	11.13 (6.02, 20.58)	< 0.001
2011	1.48 (1.45, 1.52)	< 0.001	0.65 (0.62, 0.69)	< 0.001	10.50 (5.68, 19.44)	< 0.001
2012	1.45 (1.42, 1.48)	< 0.001	0.77 (0.73, 0.81)	< 0.001	16.11 (8.78, 29.57)	< 0.001
2013	1.34 (1.31, 1.37)	< 0.001	0.75 (0.72, 0.79)	< 0.001	12.38 (6.72, 22.81)	< 0.001
2014	1.41 (1.38, 1.44)	< 0.001	0.83 (0.79, 0.87)	< 0.001	13.84 (7.53, 25.45)	< 0.001
2015	1.50 (1.47, 1.53)	< 0.001	1.12 (1.07, 1.17)	< 0.001	17.81 (9.73, 32.61)	< 0.001
2016	1.62 (1.58, 1.65)	< 0.001	1.53 (1.47, 1.60)	< 0.001	18.18 (9.93, 33.27)	< 0.001

*Models for RR estimates include log(population estimate) and county-level random effects. **Chlamydia case data is missing for the year 2000.

Table 6: Rate ratios for annual STI counts in relation to drilling and sociodemographic characteristics in North Dakota (2000-2016).

	North Dakota					
	Chlamydia		Gonorrhea		Syphilis	
Model Parameters	RR* (95% CI)	P-Value	RR* (95% CI)	P-Value	RR* (95% CI)	P-Value
Active Shale Spuds (ref=0)	1.00	--	1.00	--	1.00	--
Low (< 50)	1.02 (0.95, 1.09)	0.586	1.25 (1.00, 1.56)	0.046	2.53 (0.69, 9.26)	0.161
High (≥ 50)	1.44 (1.32, 1.56)	< 0.001	1.69 (1.30, 2.20)	< 0.001	1.69 (0.40, 7.12)	0.472
Population Density (ppsm)	0.78 (0.76, 0.80)	< 0.001	1.26 (1.19, 1.33)	< 0.001	2.93 (1.89, 4.55)	< 0.001
% Male	1.13 (1.11, 1.15)	< 0.001	2.02 (1.88, 2.15)	< 0.001	2.12 (1.50, 3.00)	< 0.001
% Female	0.88 (0.87, 0.90)	< 0.001	0.50 (0.46, 0.53)	< 0.001	0.47 (0.33, 0.66)	< 0.001
% 15-29	0.90 (0.86, 0.93)	< 0.001	0.57 (0.51, 0.63)	< 0.001	0.96 (0.65, 1.42)	0.836
% White	0.25 (0.22, 0.27)	< 0.001	0.00 (0.00, 0.00)	< 0.001	0.49 (0.36, 0.67)	< 0.001
% Black	1.05 (1.04, 1.05)	< 0.001	1.10 (1.09, 1.11)	< 0.001	1.18 (1.10, 1.26)	< 0.001
% Asian	1.03 (1.02, 1.04)	< 0.001	1.09 (1.08, 1.10)	< 0.001	1.04 (0.97, 1.12)	0.227
% Hispanic	1.20 (1.18, 1.23)	< 0.001	2.38 (2.24, 2.52)	< 0.001	2.53 (1.86, 3.43)	< 0.001
% Native	1.90 (1.65, 2.19)	< 0.001	212.77 (135.56, 333.95)	< 0.001	1.75 (1.29, 2.36)	0.001
% High School	0.79 (0.76, 0.82)	< 0.001	0.42 (0.39, 0.45)	< 0.001	0.50 (0.32, 0.77)	0.002
% Bachelor's Degree	1.33 (1.29, 1.37)	< 0.001	3.44 (3.21, 3.70)	< 0.001	4.14 (2.33, 7.36)	< 0.001
Median Household Income (USD)	1.22 (1.20, 1.23)	< 0.001	2.83 (2.67, 3.00)	< 0.001	4.60 (2.97, 7.12)	< 0.001
% Below Poverty	1.10 (1.05, 1.14)	< 0.001	1.48 (1.34, 1.63)	< 0.001	0.94 (0.64, 1.38)	0.743
% Health Insurance	1.59 (1.48, 1.70)	< 0.001	7.62 (6.63, 8.77)	< 0.001	0.74 (0.51, 1.05)	0.091
Year						
2000**	--	--	--	--	--	--
2001**	--	--	--	--	--	--
2002	1.00 (ref)	--	1.00 (ref)	--	--	--
2003	0.93 (0.86, 1.00)	0.050	1.28 (0.94, 1.73)	0.115	1.00 (ref)	--
2004**	1.00 (0.93, 1.08)	0.919	1.35 (1.00, 1.82)	0.050	--	--
2005	0.92 (0.86, 1.00)	0.039	1.57 (1.17, 2.10)	0.003	0.49 (0.04, 5.47)	0.565
2006	1.00 (0.93, 1.08)	0.944	1.86 (1.40, 2.47)	< 0.001	0.49 (0.04, 5.42)	0.560
2007	0.98 0.91, 1.06)	0.601	1.40 (1.04, 1.89)	0.025	0.49 (0.04, 5.40)	0.557
2008**	1.04 (0.97, 1.12)	0.241	1.72 (1.29, 2.29)	< 0.001	--	--
2009	1.05 (0.98, 1.13)	0.178	1.79 (1.35, 2.38)	< 0.001	1.91 (0.35, 10.46)	0.455
2010	1.24 (1.15, 1.32)	< 0.001	2.34 (1.78, 3.07)	< 0.001	1.38 (0.23, 8.30)	0.724
2011	1.23 (1.15, 1.32)	< 0.001	2.83 (2.17, 3.69)	< 0.001	0.45 (0.04, 5.00)	0.517
2012	1.43 (1.33, 1.53)	< 0.001	3.68 (2.84, 4.77)	< 0.001	1.76 (0.32, 9.63)	0.515
2013	1.39 (1.30, 1.49)	< 0.001	5.24 (4.08, 6.74)	< 0.001	5.11 (1.14, 22.91)	0.033
2014	1.60 (1.49, 1.71)	< 0.001	7.25 (5.66, 9.28)	< 0.001	5.42 (1.22, 24.09)	0.027
2015	1.43 (1.33, 1.53)	< 0.001	7.01 (5.48, 8.98)	< 0.001	4.50 (0.99, 20.37)	0.051
2016	1.56 (1.46, 1.67)	< 0.001	10.25 (8.04, 13.08)	< 0.001	13.51 (3.23, 56.45)	< 0.001

*Models for RR estimates include log(population estimate) and county-level random effects. **Chlamydia and Gonorrhea case data are missing for the years 2000 and 2001. Syphilis case data is missing for the years 2000, 2001, 2002, 2004, and 2008.

Table 7: Rate ratios for annual STI counts in relation to drilling and sociodemographic characteristics in Texas (2000-2016).

	Texas					
	Chlamydia		Gonorrhea		Syphilis	
Model Parameters	RR* (95% CI)	P-Value	RR* (95% CI)	P-Value	RR* (95% CI)	P-Value
Active Shale Spuds (ref=0)	1.00	--	1.00	--	1.00	--
Low (< 50)	1.21 (1.20, 1.22)	< 0.001	0.97 (0.96, 0.99)	< 0.001	0.95 (0.89, 1.00)	0.066
High (≥ 50)	1.21 (1.20, 1.23)	< 0.001	0.85 (0.83, 0.87)	< 0.001	1.16 (1.04, 1.30)	0.010
Population Density (ppsm)	1.47 (1.47, 1.48)	< 0.001	1.06 (1.06, 1.07)	< 0.001	1.61 (1.57, 1.66)	< 0.001
% Male	0.77 (0.76, 0.78)	< 0.001	1.00 (0.98, 1.02)	0.790	0.83 (0.75, 0.92)	< 0.001
% Female	1.30 (1.28, 1.32)	< 0.001	1.00 (0.98, 1.03)	0.790	1.20 (1.09, 1.33)	< 0.001
% 15-29	0.59 (0.59, 0.60)	< 0.001	0.93 (0.91, 0.94)	< 0.001	0.39 (0.36, 0.42)	< 0.001
% White	1.24 (1.24, 1.25)	< 0.001	1.11 (1.10, 1.12)	< 0.001	1.84 (1.75, 1.94)	< 0.001
% Black	2.27 (2.24, 2.31)	< 0.001	1.16 (1.13, 1.20)	< 0.001	1.86 (1.70, 2.03)	< 0.001
% Asian	1.28 (1.28, 1.28)	< 0.001	1.09 (1.09, 1.10)	< 0.001	**	**
% Hispanic	4.01 (3.95, 4.06)	< 0.001	1.14 (1.11, 1.16)	< 0.001	6.61 (5.90, 7.41)	< 0.001
% Native	1.00 (1.00, 1.01)	0.089	1.02 (1.01, 1.03)	< 0.001	0.92 (0.86, 0.99)	0.022
% High School	1.19 (1.18, 1.20)	< 0.001	0.94 (0.92, 0.95)	< 0.001	1.37 (1.26, 1.49)	< 0.001
% Bachelor’s Degree	2.00 (1.98, 2.01)	< 0.001	1.26 (1.24, 1.28)	< 0.001	3.77 (3.47, 4.08)	< 0.001
Median Household Income (USD)	1.49 (1.48, 1.49)	< 0.001	1.11 (1.11, 1.12)	< 0.001	2.15 (2.07, 2.23)	< 0.001
% Below Poverty	1.63 (1.61, 1.64)	< 0.001	0.96 (0.94, 0.97)	< 0.001	2.10 (1.97, 2.23)	< 0.001
% Health Insurance	1.47 (1.46, 1.48)	< 0.001	1.25 (1.24, 1.27)	< 0.001	1.75 (1.66, 1.84)	< 0.001
Year						
2000	1.00 (ref)	--	1.00 (ref)	--	1.00 (ref)	--
2001	1.00 (0.99, 1.01)	0.580	0.90 (0.88, 0.91)	< 0.001	1.19 (1.04, 1.36)	0.012
2002	0.98 (0.97, 0.99)	< 0.001	0.80 (0.78, 0.81)	< 0.001	1.44 (1.27, 1.64)	< 0.001
2003	0.96 (0.95, 0.97)	< 0.001	0.72 (0.70, 0.73)	< 0.001	1.58 (1.39, 1.79)	< 0.001
2004	0.96 (0.95, 0.97)	< 0.001	0.70 (0.70, 0.71)	< 0.001	1.97 (1.75, 2.22)	< 0.001
2005	0.97 (0.96, 0.98)	< 0.001	0.74 (0.73, 0.75)	< 0.001	2.05 (1.82, 2.31)	< 0.001
2006	1.00 (0.99, 1.01)	0.380	0.84 (0.83, 0.86)	< 0.001	2.44 (2.17, 2.73)	< 0.001
2007	1.11 (1.10, 1.12)	< 0.001	0.87 (0.86, 0.89)	< 0.001	2.61 (2.32, 2.92)	< 0.001
2008	1.28 (1.27, 1.29)	< 0.001	0.86 (0.85, 0.87)	< 0.001	3.10 (2.77, 3.46)	< 0.001
2009	1.32 (1.31, 1.33)	< 0.001	0.77 (0.76, 0.78)	< 0.001	3.55 (3.18, 3.97)	< 0.001
2010	1.46 (1.45, 1.48)	< 0.001	0.82 (0.81, 0.83)	< 0.001	2.63 (2.34, 2.94)	< 0.001
2011	1.50 (1.49, 1.52)	< 0.001	0.79 (0.77, 0.80)	< 0.001	2.45 (2.19, 2.75)	< 0.001
2012	1.50 (1.49, 1.51)	< 0.001	0.81 (0.80, 0.82)	< 0.001	3.35 (3.00, 3.74)	< 0.001
2013	1.51 (1.50, 1.52)	< 0.001	0.83 (0.82, 0.85)	< 0.001	2.99 (2.68, 3.34)	< 0.001
2014	1.50 (1.49, 1.51)	< 0.001	0.85 (0.84, 0.87)	< 0.001	3.26 (2.92, 3.63)	< 0.001
2015	1.59 (1.57, 1.60)	< 0.001	0.95 (0.93, 0.96)	< 0.001	3.28 (2.94, 3.66)	< 0.001
2016	1.60 (1.59, 1.62)	< 0.001	1.01 (0.99, 1.02)	0.231	3.82 (3.43, .26)	< 0.001

*Models for RR estimates include log(population estimate) and county-level random effects. **Model does not converge for association between % Asian and Syphilis.

Table 8: Spearman correlation of standardized sociodemographic covariates in Colorado (2000-2016).

	Population Density (ppsm)	% Male	% Female	% 15-29	% White	% Black	% Asian	% Hispanic	% Native	% High School	% Bachelor's Degree	% Poverty	% Health Insurance	% Med. Household Income (USD)
Population Density (ppsm)	1.000													
% Male	-0.110	1.000												
% Female	0.110	-1.000	1.000											
% 15-29	0.372	0.193	-0.193	1.000										
% White	-0.429	0.074	-0.074	-0.416	1.000									
% Black	0.447	0.132	-0.132	0.251	-0.452	1.000								
% Asian	0.531	0.036	-0.036	0.357	-0.270	0.362	1.000							
% Hispanic	0.211	-0.127	0.127	0.365	-0.777	0.296	0.123	1.000						
% Native	-0.075	0.028	-0.028	0.064	-0.384	0.099	-0.107	0.259	1.000					
% High School	-0.489	-0.014	0.014	-0.255	-0.005	-0.040	-0.399	0.137	0.238	1.000				
% Bachelor's Degree	0.385	0.067	-0.067	0.122	0.281	-0.024	0.342	-0.407	-0.352	-0.820	1.000			
% Poverty	-0.197	-0.055	0.055	0.123	-0.483	0.072	-0.108	0.619	0.439	0.360	-0.556	1.000		
% Health Insurance	0.245	-0.109	0.109	-0.118	0.169	0.267	0.220	-0.304	-0.221	-0.132	0.211	-0.297	1.000	
Med. Household Income (USD)	0.470	0.076	-0.076	0.114	0.271	0.110	0.470	-0.351	-0.379	-0.552	0.722	-0.679	0.313	1.000

Table 9: Spearman correlation of standardized sociodemographic covariates in North Dakota (2000-2016).

	Population Density (ppsm)	% Male	% Female	% 15-29	% White	% Black	% Asian	% Hispanic	% Native	% High School	% Bachelor's Degree	% Poverty	% Health Insurance	% Med. Household Income (USD)
Population Density (ppsm)	1.000													
% Male	-0.037	1.000												
% Female	0.037	-1.000	1.000											
% 15-29	0.627	0.087	-0.087	1.000										
% White	-0.450	-0.040	0.040	-0.610	1.000									
% Black	0.384	0.102	-0.102	0.394	-0.416	1.000								
% Asian	0.391	0.042	-0.042	0.380	-0.384	0.404	1.000							
% Hispanic	0.388	0.195	-0.195	0.424	-0.465	0.360	0.281	1.000						
% Native	0.301	-0.015	0.015	0.460	-0.839	0.265	0.273	0.266	1.000					
% High School	-0.476	0.067	-0.067	-0.420	0.361	-0.217	-0.141	-0.229	-0.302	1.000				
% Bachelor's Degree	0.304	0.085	-0.085	0.403	-0.145	0.371	0.280	0.267	0.004	-0.298	1.000			
% Poverty	-0.077	-0.021	0.021	0.052	-0.141	-0.117	-0.089	-0.057	0.194	-0.006	-0.409	1.000		
% Health Insurance	0.034	-0.070	0.070	-0.245	0.438	0.036	-0.101	-0.054	-0.439	-0.004	0.132	-0.387	1.000	
Med. Household Income (USD)	0.117	0.327	-0.327	0.135	-0.236	0.389	0.296	0.401	0.123	-0.042	0.550	-0.576	0.149	1.000

Table 10: Spearman correlation of standardized sociodemographic covariates in Texas (2000-2016).

	Population Density (ppsm)	% Male	% Female	% 15-29	% White	% Black	% Asian	% Hispanic	% Native	% High School	% Bachelor's Degree	% Poverty	% Health Insurance	% Med. Household Income (USD)
Population Density (ppsm)	1.000													
% Male	-0.092	1.000												
% Female	0.092	-1.000	1.000											
% 15-29	0.351	0.234	-0.234	1.000										
% White	-0.309	-0.052	0.052	-0.444	1.0000									
% Black	0.488	0.052	-0.052	0.231	-0.621	1.000								
% Asian	0.521	-0.016	0.016	0.307	-0.182	0.273	1.000							
% Hispanic	-0.169	0.182	-0.182	0.382	-0.173	-0.320	-0.024	1.000						
% Native	0.041	0.024	-0.024	0.038	0.022	-0.065	0.035	-0.027	1.000					
% High School	-0.156	0.112	-0.112	-0.327	0.069	0.255	-0.227	-0.413	-0.016	1.000				
% Bachelor's Degree	0.188	-0.144	0.144	-0.102	0.225	-0.147	0.275	-0.161	0.021	-0.411	1.000			
% Poverty	-0.143	-0.021	0.021	0.237	-0.244	0.051	-0.069	0.407	-0.071	-0.018	-0.408	1.000		
% Health Insurance	0.147	0.013	-0.013	-0.167	0.047	0.195	0.163	-0.376	-0.029	0.042	0.385	-0.465	1.000	
Med. Household Income (USD)	0.302	0.137	-0.137	-0.026	0.170	0.025	0.284	-0.117	0.064	-0.077	0.522	-0.608	0.410	1.000

Table 11: Rate ratios (RR) and 95% confidence intervals (95% CI) for the association between shale activity and sexually transmitted infections (STI) in Colorado, North Dakota, and Texas (2000-2016).

	Colorado			North Dakota			Texas		
	County Years* (N=1,088)	Unadjusted RR** (95% CI)	Adjusted RR[A] (95% CI)	County Years* (N=901)	Unadjusted RR** (95% CI)	Adjusted RR[B] (95% CI)	County Years* (N=4,318)	Unadjusted RR** (95% CI)	Adjusted RR[C] (95% CI)
Chlamydia									
No Shale Spuds (ref)	880	1.00	--	691	1.00	--	3,696	1.00	--
< 50 Shale Spuds	127	1.07 (0.96, 1.19)	1.01 (0.94, 1.07)	61	1.31 (1.00, 1.71)	0.89 (0.72, 1.11)	481	1.14 (1.09, 1.19)	1.02 (0.99, 1.06)
≥ 50 Shale Spuds	17	0.86 (0.64,1.15)	0.82 (0.69, 0.99)	43	1.88 (1.38, 2.55)	0.96 (0.73, 1.27)	141	1.36 (1.26, 1.47)	1.10 (1.04, 1.17)
Gonorrhea									
No Shale Spuds (ref)	938	1.00	--	691	1.00	--	3,696	1.00	--
< 50 Shale Spuds	133	1.00 (0.83, 1.20)	1.06 (0.91, 1.22)	61	1.31 (0.76, 2.27)	1.04 (0.67, 1.62)	481	1.02 (0.96, 1.08)	1.02 (0.96, 1.08)
≥ 50 Shale Spuds	17	1.03 (0.64, 1.67)	1.01 (0.70, 1.45)	43	2.25 (1.21, 4.19)	1.12 (0.64, 1.95)	141	1.13 (1.02, 1.26)	1.15 (1.04, 1.28)
Syphilis									
No Shale Spuds (ref)	938	1.00	--	544	1.00	--	3,696	1.00	--
< 50 Shale Spuds	133	1.74 (1.19, 2.55)	1.21 (0.93, 1.57)	52	2.03 (0.52, 7.99)	0.93 (0.14, 6.37)	481	1.23 (0.91, 1.67)	0.94 (NA, NA)
≥ 50 Shale Spuds	17	4.06 (1.61, 10.26)	1.33 (0.65, 2.73)	40	1.62 (0.38, 6.94)	0.63 (0.09, 4.25)	141	1.33 (0.79, 2.25)	0.88 (NA, NA)

*County years may not sum to total due to missing data.

**Unadjusted models include log(population) offset, county-level random effects, and observation-level random effects.

[A] **Colorado** – Adjusted model includes log(population) offset, county-level random effects, observation-level random effects, and adjustment for population density (ppsm), % female, % 15-29, % black, % Hispanic, % high school, % poverty, % health insurance, and year.

[B] **North Dakota** – Adjusted model includes log(population) offset, county-level random effects, observation-level random effects, and adjustment for population density (ppsm), % female, % 15-29, % white % black, % Hispanic, % American Indian or Alaska Native , % high school, % Bachelor's degree, % poverty, % health insurance, median household income (USD), and year.

[C] **Texas** – Adjusted model includes log(population) offset, county-level random effects, observation-level random effects, and adjustment for population density (ppsm), % female, % 15-29, % white, % black, % Hispanic, % high school, % Bachelor's degree, % poverty, % health insurance, median household income (USD), and year. 95% CI unavailable for syphilis model in Texas.